**Assignment-02**

Problem 1

Loading the file and assigning the Muscle Mass as Y and Age as X

Data=read.table("AS\_2\_Q\_1\_Data.txt", header = FALSE, sep = "") Y = Data[1]$V1

X = Data[2]$V2

Creating a linear model and assigning the values of estimators.

Model=lm(Y~X) Model

##

## Call:

## lm(formula = Y ~ X) ##

## Coefficients:

## (Intercept) X

## 156.35 -1.19

Beta0=as.numeric(Model$coefficients["(Intercept)"]) Beta1=as.numeric(Model$coefficients["X"])

1.a.

Hypothesis Testing

Let’s Start a two sided Null Hypothesis Testing Assumption: H0: Beta1 is a Positive Integer Beta1>=0 Ha: Beta1 is a Negative Integer Beta1<0

Decision: As we can see the P value is 4.12\*10ˆ-19 which is far less than alpha=0.05. So we reject the null hypothesis.

Conclusion: Beta1 is a Negative Integer Beta1<0 The P value of the two sided test is 4.12\*10ˆ-19

SXY = sum(X \* Y) - length(X) \* mean(X) \* mean(Y) SYY = sum(Y \* Y) - length(Y) \* (mean(Y))ˆ2

SXX = sum(X \* X) - length(X) \* (mean(X))ˆ2

RSS = SYY - Beta1 \* SXY SIGMA=(RSS/(length(X)-2))ˆ0.5

Tobs=Beta1\*(SXX)ˆ0.5/SIGMA Tobs

## [1] -13.19326

2\*pt(q=Tobs, df=58,lower.tail=TRUE)

## [1] 4.123987e-19

1.b.

No even the test of Non Zero Beta0 is significant. We can not provide information on amount of muscle mass at birth of a female child because we do not have data collected in that range.

1.c.

The 95 % confidence interval of muscle mass for the woman whose ages differ by one year is Beta1(The Slope of the Linear Model). Calculating the confidence interval gives a range (-1.370545,-1.009446)

It is not necessary to know the specific ages because the difference in muscle mass of women whose age is differed by one year is Beta1 which is the slope of the regression model.

n=0.05

t=qt(n/2, length(X)-2, lower.tail=TRUE)

lb=Beta1+t\*(SIGMA/SXXˆ0.5) ub=Beta1-t\*(SIGMA/SXXˆ0.5) lb

## [1] -1.370545

ub

## [1] -1.009446

2.a.

The values of Beta0 and Beta1 are 2.114,0.038 The model is #Y=2.11+0.03883X

Data =read.delim("AS\_2\_Q\_2\_Data.txt", header = FALSE, sep = " ") Y=Data[,2]

X=Data[,6]

Model=lm(Y~X)

Beta0=as.numeric(Model$coefficients["(Intercept)"]) Beta0

## [1] 2.114049

Beta1=as.numeric(Model$coefficients["X"]) Beta1

## [1] 0.03882713

2.b.

The plot of the Data and the Model Fit is as shown. The regression function fit the model well with minimum outliers.

plot(X,Y,xlab="ACT Test Score",ylab="Average GPA") abline(Beta0,Beta1,col=2,lwd=2)

2.5

3.5

# 15 20 25 30 35



Average GPA

0.5

1.5

ACT Test Score

2.c.

The point estimate of mean freshman GPA of ACT score=30 is 3.2788

Y1=Beta0+Beta1\*30 Y1

## [1] 3.278863

2.d.

The point estimate of change in mean response when entrance score is increased by one point is Beta1 which is 0.0388

Problem 3 3.a.

The 95 % confidence interval of test score which is 28 is (3.06,3.34)

xnew = data.frame(X=28) Y2=Beta0+Beta1\*28

SXY = sum(X \*Y ) - length(X) \* mean(X) \* mean(Y) SYY = sum( Y \* Y) - length(Y) \* (mean(Y))ˆ2

SXX = sum( X \* X) - length(X) \* (mean(X))ˆ2 RSS = SYY - Beta1 \* SXY SIGMA=(RSS/(length(X)-2))ˆ0.5

mu=mean(X)

n=0.05

t=qt(n/2, length(X)-2, lower.tail=TRUE)

*#Method 1 from Function*

confInt = predict(Model,xnew, interval = "confidence", level = 0.95, se.fit = TRUE) conflwr=confInt$fit[2]

confupr=confInt$fit[3] conflwr

## [1] 3.061384

confupr

## [1] 3.341033

*#Method 2 from Formula* ConfUb=Y2+t\*SIGMA\*((1/length(X))+((xnew-mu)ˆ2/SXX))ˆ0.5 ConfLb=Y2-t\*SIGMA\*((1/length(X))+((xnew-mu)ˆ2/SXX))ˆ0.5 ConfUb

## X

## 1 3.061384

ConfLb

## X

## 1 3.341033

3.b.

The 95 % prediction interval of test score which is 28 is (1.95,4.44)

*#Method 1 from Function*

predInt=predict(Model,xnew, interval = "prediction", level = 0.95, se.fit = TRUE) predlwr=predInt$fit[2]

predlwr

## [1] 1.959355

predupr=predInt$fit[3] predupr

## [1] 4.443063

*#Method 2 Derivating with Formula* PredUb=Y2+t\*SIGMA\*(1+(1/length(X))+((xnew-mu)ˆ2/SXX))ˆ0.5 PredLb=Y2-t\*SIGMA\*(1+(1/length(X))+((xnew-mu)ˆ2/SXX))ˆ0.5 PredUb

## X

## 1 1.959355

PredLb

## X

## 1 4.443063

3.c.

Yes the Prediction interval is wider than confidence interval. Because the prediction model is used to find estimates of the random samples rather than the confidence interval which is an inference on sample data.

3.d.

The range of confidence band is (3.02,3.37).

Yes the confidence band is little wider because it represents the entire regression model line not only the sample at Xh=28.

Weight = sqrt(2\*qf(0.95,2,length(X)-2))

cbandupper = confInt$fit[,1]+Weight\*confInt$se.fit cbandlower = confInt$fit[,1]-Weight\*confInt$se.fit cbandlower

## [1] 3.026159

cbandupper

## [1] 3.376258

4.a.

Anova Table

Anova =anova(Model) Anova

## Analysis of Variance Table ##

## Response: Y

## Df Sum Sq Mean Sq F value Pr(>F)

## X 1 3.588 3.5878 9.2402 0.002917 \*\*

## Residuals 118 45.818 0.3883 ## ---

## Signif. codes: 0 ’\*\*\*’ 0.001 ’\*\*’ 0.01 ’\*’ 0.05 ’.’ 0.1 ’ ’ 1

4.b.

MSR is the sum of squares due to regression taking degrees of freedom into account.

MSE is Mean Squared Error which is not a biased estimator of Standard Deviation (Variance Squared). When Beta1=0 or the slope of the regression equation is 0.

4.c.

Let’s Start Null Hypothesis Testing

Assumption: H0: Beta1 is zero Beta1=0 Ha: Beta1 is not Zero Beta1!=0

Decision rule: The value of Fscore is less than FStarScore so we reject the null hypothesis so the Alternative is true which is Beta1 is not equal to 0.

Conclusion: Beta1 is not Zero Beta1!=0

alpha=0.01

Fscore = qf((1-alpha),1,length(X)-2) Fscore

## [1] 6.854641

FStarScore=9.242 *#From Anova Table*

FStarScore

## [1] 9.242

4.d.

The absolute magnitude of reduction in variance of Y when X is introduced in regression model is R Squared(Rˆ2)

4.e.

The sign of R is positive as the data has a positive correlation from the graph.

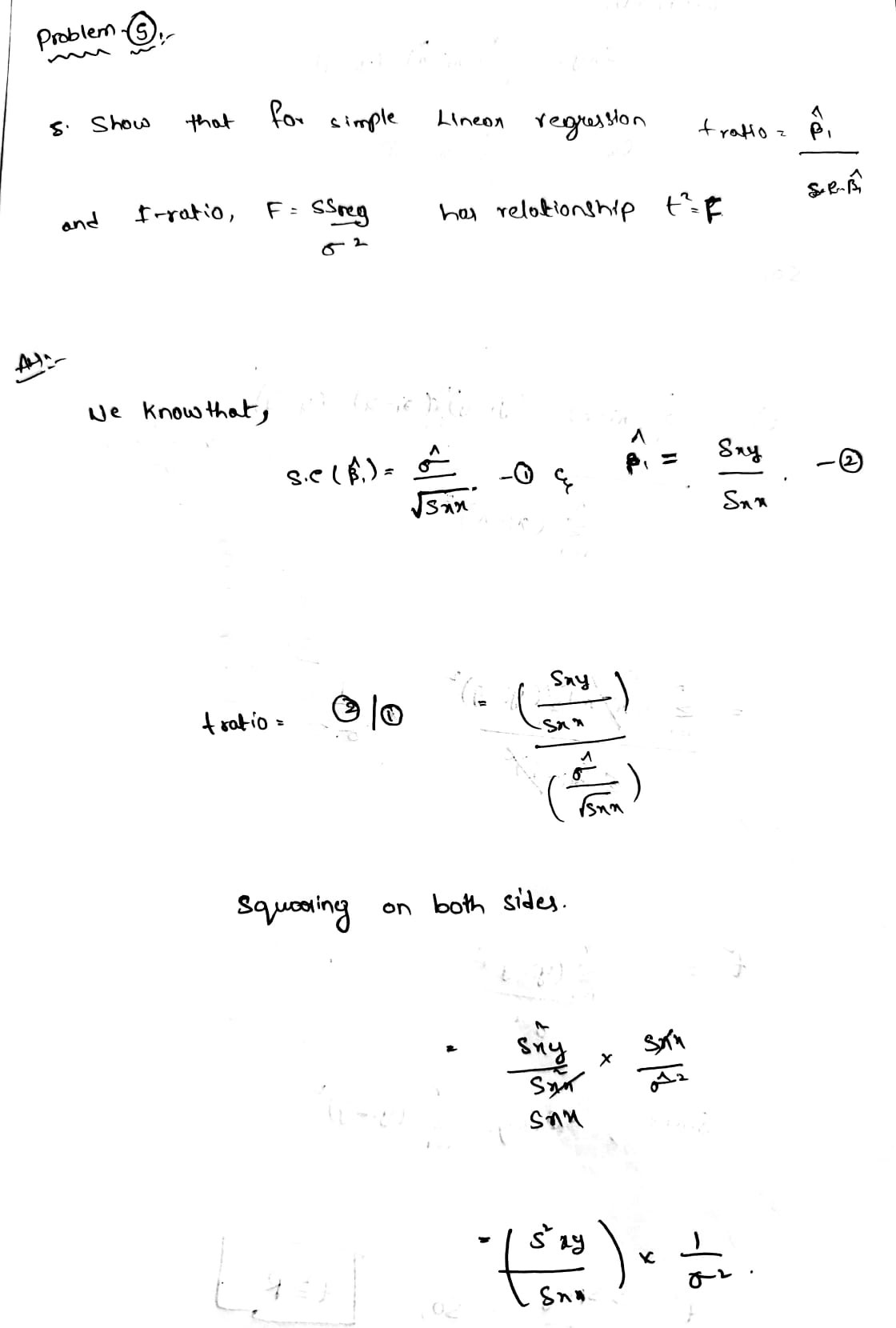
R=sqrt(0.07262) *# from summary*

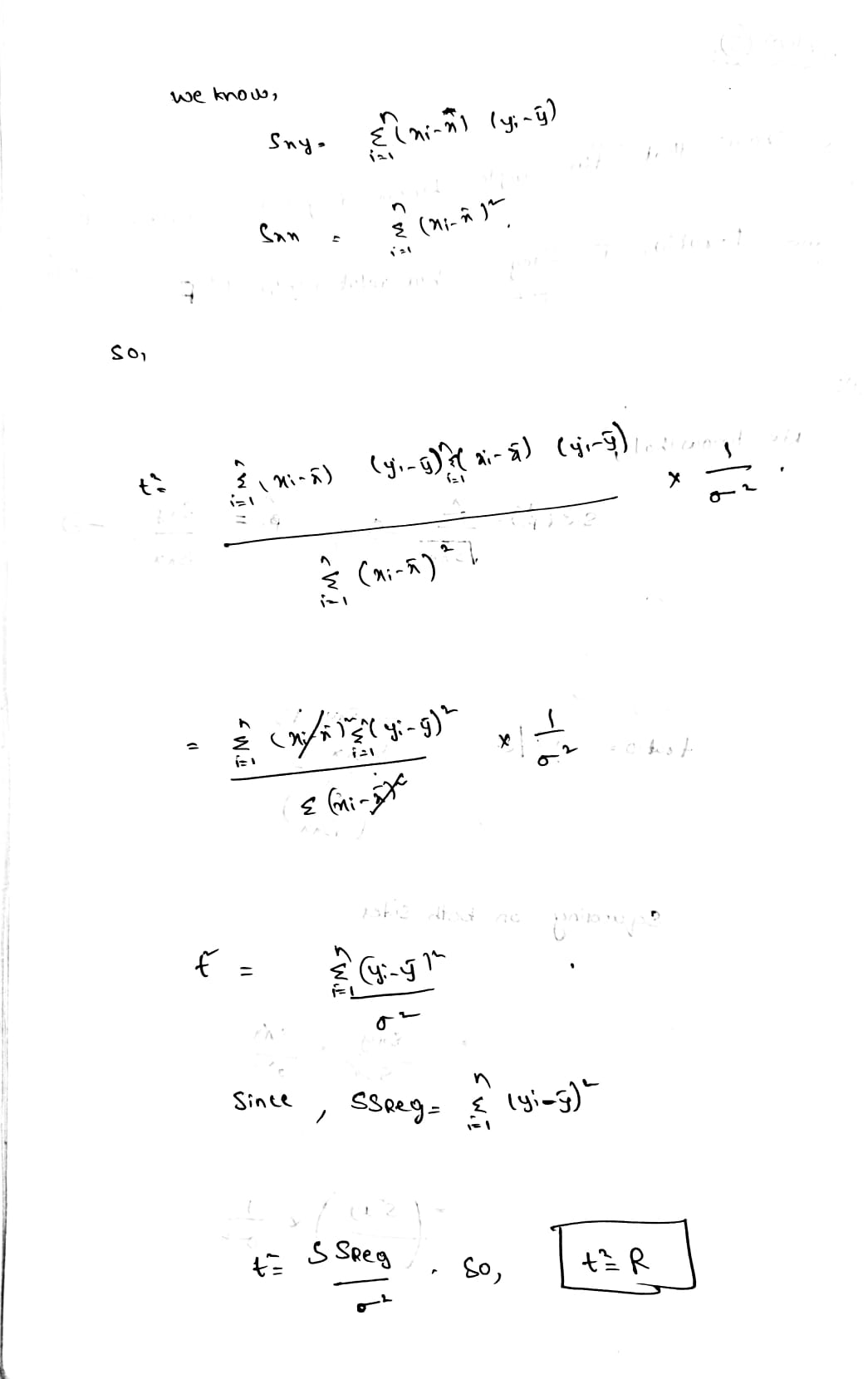
R

## [1] 0.269481

4.f.

R Squared gives clear cut interpretation. It describes the percent of variance of Y with respect to X. R square is usually used to represent the relation. It takes values from 0 to 1.

****

****